THE HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY

Department of Mathematics

MATHEMATICS COLLOQUIUM

Interaction of Statistics and Geometry: A New Landscape for Data Science

By

Prof. Zhigang YAO
National University of Singapore

Abstract

While classical statistics primarily focuses on observations as real numbers or elements of real vector spaces, contemporary statistical challenges often involve more complex data types. These data are represented in spaces that, although not strictly Euclidean vector spaces, possess inherent geometric structures. The community exploring the interaction between statistics and geometry is expanding in both numbers and scope. The concept of manifold fitting traces back to H. Whitney's work in the early 1930s. The resolution of the Whitney extension problem has yielded new insights into data interpolation and inspired the formulation of the Geometric Whitney Problems. Specifically, given a set, we inquire: when can we construct a smooth d-dimensional submanifold to approximate the set, and how effectively can we estimate it in terms of distance and smoothness? In this talk, I will explore the manifold fitting problem, highlighting its modern insights and implications. Although various mathematical approaches have been proposed, many rely on restrictive assumptions, complicating the development of efficient and practical algorithms. As the manifold hypothesis-exploring non-Euclidean structures-remains a cornerstone of data science, further exploration of the manifold fitting problem is essential within the contemporary data science community. This discussion will be informed by recent work by Yao, Yau, and other co-authors, alongside ongoing research.

Bio: Zhigang Yao is a tenured Associate Professor in the Department of Statistics and Data Science at the National University of Singapore. Since 2022, he has also been a visiting faculty member at the Center for Mathematical Sciences and Applications at Harvard University. In addition, he holds visiting professorships at the YMSC at Tsinghua University and the Shanghai Institute of Mathematical Sciences and Interdisciplinary Science (SIMIS).Yao's primary research interests are in statistical inference for complex data. In recent years, his focus has shifted towards Non-Euclidean Statistics and low-dimensional manifold fitting. He is dedicated to advancing the emerging field at the intersection of geometry and statistics. Along with his collaborators, Yao has proposed novel methods and theories that extend traditional principal component analysis (PCA) to Riemannian manifolds, including principal flows, sub-manifolds, and principal boundaries. These innovations offer new manifold fitting theories designed to address the limitations of conventional statistical methods by incorporating the geometric structure of data.

Date   :   **11 April 2025 (Friday)**
Time   :   **3:00p.m.-4:00p.m.**
Venue   :   **Lecture Theatre F (Lift 25/26)**